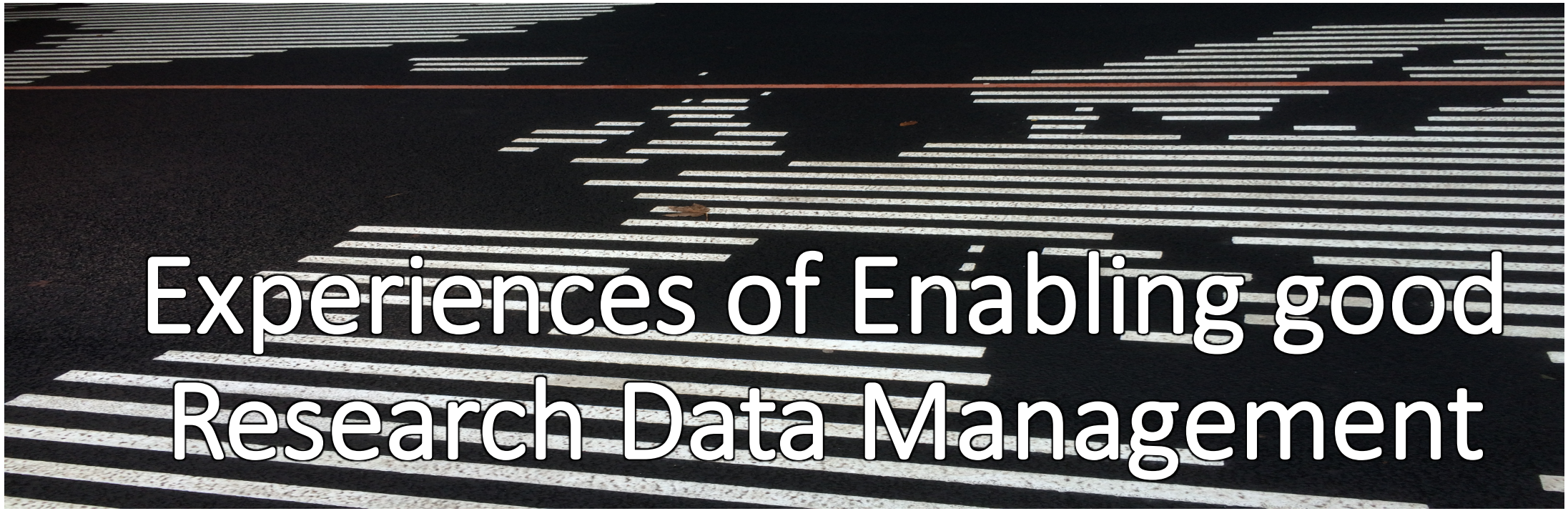
Experiences of Enabling good Research Data Management Practice



Experiences of Enabling good Research Data Management

Stuart Lewis

Deputy Director, Library & University Collections

@stuartlewis

THE UNIVERSITY of EDINBURGH

Opin vísindi. Hvað verður um rannsóknargögnin þín?

Open Science – what will become of your research data?

1. The context – The University of Edinburgh
2. The theory – Reasons for Research Data Management
3. The policy – Local and national policies
4. The practice – An overview of the Research Data Service

THE UNIVERSITY *of* EDINBURGH

What is your role?

Poll locked. Responses not accepted.

“Librarian” about 3 hours ago	“Research/admin” about 3 hours ago	“Research support” about 3 hours ago	“Academic” about 3 hours ago	“Research” about 3 hours ago	“Academic” about 3 hours ago
“Academic” about 3 hours ago	“Public administration” about 3 hours ago	“Librarian” about 3 hours ago	“Senior advisor at the Ministry” about 3 hours ago	“Yes” about 3 hours ago	
“researcher in humanities” about 3 hours ago	“librarian” about 3 hours ago	“Research support / admin.” about 3 hours ago	“Librarian” about 3 hours ago	“Professor” about 3 hours ago	
“academic” about 3 hours ago	“Directo of library and inf services” about 3 hours ago	“Research” about 3 hours ago	“Research director” about 3 hours ago		
“Research administration” about 3 hours ago	“Project manager department of science” about 3 hours ago	“IT and librarian” about 3 hours ago			

The context

The University of Edinburgh

- The context to our work:
  - Edinburgh is the capital city of Scotland
  - The University of Edinburgh is Scotland's largest university
  - A large thriving research-led University: 35,258 students, 13,272 staff
  - Breadth of research disciplines across three colleges:
    - Humanities and Social Science
    - Science and Engineering
    - Medicine and Veterinary Medicine

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES LERU

RUSSELL PIONEERING RESEARCH GROUP

UNIVERSITAS 21

UNIVERSITIES COIMBRA GROUP

THE UNIVERSITY of EDINBURGH

The University of Edinburgh

- Information Services Group at the University of Edinburgh - CIO
  - Library & University Collections
  - User Services
  - IT Infrastructure
  - IT Applications
  - Learning Teaching and Web

  - Digital Curation Centre and EDINA

The University of Edinburgh

- The Research Data Service (SO + SOM)
  - Library & University Collections
  - User Services
  - IT Infrastructure
  - IT Applications
  - Learning Teaching and Web

  - Digital Curation Centre and EDINA






THE UNIVERSITY *of* EDINBURGH

The Theory

What is research data?

- Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings

  https://www.whitehouse.gov/omb/circulars_a110#36

- Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship

  Christine Borgman, UCLA, 2014: *Big Data, Little Data, No Data: Scholarship in the Networked World.* MIT Press

Good collection of definitions:
http://www2.le.ac.uk/services/research-data/documents/UoL_ReserchDataDefinitions_20120904.pdf

THE UNIVERSITY *of* EDINBURGH

What is research data?

- Observational: real time, unique and irreplaceable
- Experimental: reproducible, may be expensive
- Simulation: modeled data; model & metadata may be more important than output data
- Derived or compiled: combining 'raw' data, reproducible, may be expensive
- Reference or canonical: collection of peer reviewed datasets, published and curated

Research Information Network. "Stewardship of digital research data - principles and guidelines http://www.rin.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf

- Not just 'scientific' data!

THE UNIVERSITY *of* EDINBURGH

Why manage research data?

- It's nothing new – it's just **good practice**!

- It's **part of our mission** - The University's mission is the creation, dissemination and curation of knowledge

- "Monash University recognises **significant value** in the data generated by its large investment in research"

THE UNIVERSITY *of* EDINBURGH

Why manage research data?

- Johns Hopkins Univeristy: “To **protect researchers and the university**.”

- University of Tennesee: “**Protect** the faculty’s and University’s **property rights** by addressing definition, responsibility, control, and distribution of Research Data produced during activities supported by the University.

- “The University of Northampton recognises that **good research is underpinned by good research data management**.”

THE UNIVERSITY *of* EDINBURGH

Why manage research data?

- Benefits to different groups…

Seeking win + win + win + win + win……

Where do I safely keep my data from my fieldwork, as I travel home?

How can I best keep years worth of research data secure and accessible for when I and others need to re-use it?

How do we ensure we have access to our research data after some of the team have left?

How do we ensure compliance to funders' requirement for several years of open access to data?

How can our research collaborations share data, and make them available once complete?

PhD student

individual researcher

research team

university

supra-university

LEVEL

Research Integrity, London - Sept 2011

Professor Jeff Haywood

Vice Principal for Digital Education (Ex CIO)

Why manage research data?

- Benefits to different groups…

- If for no other reason, access to **my own** data!

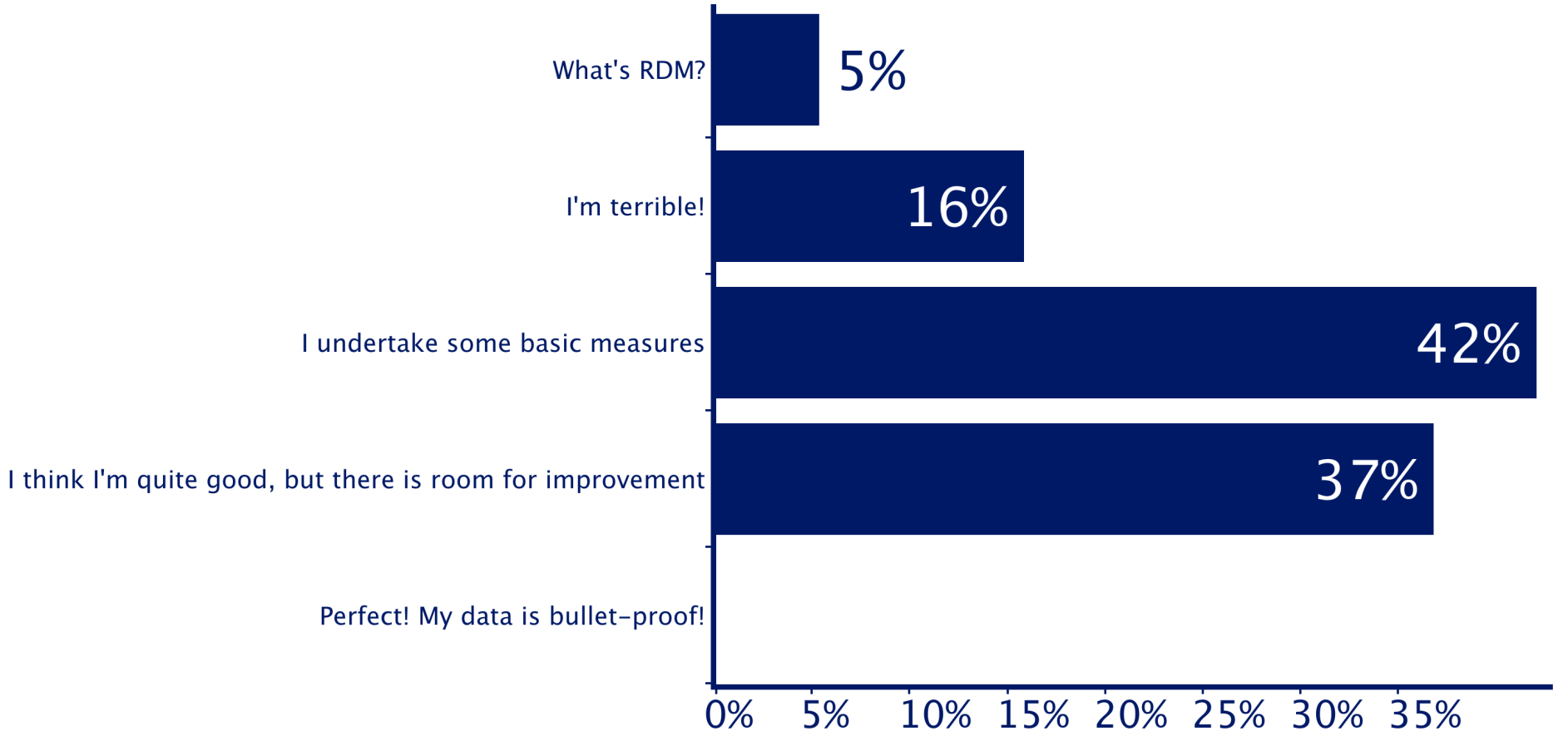
  A true story….

THE UNIVERSITY *of* EDINBURGH


LOST/STOLEN BAG
EASTPAK
Have you seen this bag?
It was stolen from here on Thursday 22nd
October around midday
It has half of my university dissertation research
inside!!
If found please call on
Reward if found. Thanks ☺

How good is your own Research Data Management practice?

Respond at **PollEv.com/rdmpoll**



The Policy

Research Data Management Policies

- Growing policy support for Research Data Management
  - University of Edinburgh Policy – May 2011
  - UK RCUK Research Councils
  - EC Horizon 2020 pilot
  - Many other research funders

THE UNIVERSITY *of* EDINBURGH

Research Data Management Policies

- University of Edinburgh Policy
  - First institutional RDM policy in the UK
    - Approved by Senatus Academicus in May 2011
    - http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy
  - An aspirational policy
    - "It is acknowledged that this is an aspirational policy, and that implementation will take some years."
  - Assigns responsibilites

THE UNIVERSITY *of* EDINBURGH

Research Data Management Policies

1. **Research data will be managed to the highest standards** throughout the research data lifecycle as part of the University's commitment to research excellence.
2. **Responsibility** for research data management through a sound research data management plan during any research project or programme **lies primarily with Principal Investigators** (PIs).
3. **All new research proposals must include research data management plans** or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. **The University** will **provide training, support, advice** and where appropriate guidelines and templates for the research data management and research data management plans.
5. **The University** will provide mechanisms and services for **storage, backup, registration, deposit and retention** of research data assets in support of current and future access, during and after completion of research projects.

THE UNIVERSITY *of* EDINBURGH

Research Data Management Policies

6. Any **data which is retained elsewhere**, for example in an international data service or domain repository **should be registered with the University**.
7. Research data management **plans** must ensure that research data are **available for access and re-use where appropriate** and under appropriate safeguards.
8. **The legitimate interests of the subjects of research data must be protected.**
9. **Research data of future historical interest**, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
10. **Exclusive rights to reuse or publish research data should not be handed over to commercial publishers** or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

THE UNIVERSITY *of* EDINBURGH

Research Councils UK (RCUK) Principles

1. Data made freely and openly available as soon as possible
2. Data Management Plans required, preservation of long-term data is required
3. Appropriate metadata should be made openly available
4. Acknowledgement of constraints on data release
5. Dataset users should acknowledge sources
6. Limited period of privileged access
7. It is appropriate to use public funds to support preservation and management of data

THE UNIVERSITY *of* EDINBURGH

UK Funder policies

● Full Coverage ◐ Partial Coverage ○ No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

THE UNIVERSITY of EDINBURGH

EPSRC Expectations

- Research organisations will promote internal awareness of these principles and expectations
- Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed
- Each research organisation will have specific policies and associated processes to maintain effective internal awareness of their publicly-funded research data holdings and of requests by third parties to access such data
- Publicly-funded research data that is not generated in digital format will be stored in a manner to facilitate it being shared
- https://www.epsrc.ac.uk/about/standards/researchdata/expectations/

THE UNIVERSITY *of* EDINBURGH

EPSRC Expectations

- Where the research data referred to in the metadata is a digital object, the metadata will include use of a robust digital object identifier
- Where access to the data is restricted the metadata should also give the reason and summarise the conditions which must be satisfied for access
- Research organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10 years from the last date on which access to the data was requested by a third party
- Research organisations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated)

THE UNIVERSITY *of* EDINBURGH

Horizon 2020

All project proposals submitted to "Research and Innovation actions" as well as "Innovation actions" include a section on research data management which is evaluated under the criterion 'Impact'. Where relevant, applicants must provide a short, general outline of their policy for data management, including the following issues:

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved?

THE UNIVERSITY *of* EDINBURGH

Horizon 2020

Regarding the digital research data generated in the action ('data'), the beneficiaries must:

a) **deposit in a research data repository** and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

- **i.** **the data**, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
- **ii.** **other data**, including associated metadata, as specified and within the deadlines laid down in the data management plan;

b) **provide information** — via the repository — about tools and instruments at thedisposal of the beneficiaries and necessary for validating the results (and —where possible — provide the tools and instruments themselves).

Link to Open Access Policies

RCUK Open Access Policy:

1. Papers must be made Open Access (within 6 or 12 months)

2. Papers must include details of the funding that supported the research

3. **If applicable, papers must include a statement on how the underlying research materials – such as data, samples or models – can be accessed**

THE UNIVERSITY *of* EDINBURGH

SECTION II.

The Practice

The University of Edinburgh's approach

- **Research Data Management services for everybody**

  We must provide core infrastructure for all researchers
  to support the transition to Open Science

THE UNIVERSITY *of* EDINBURGH

University of Edinburgh RDM Programme

- Planning a Research Data Management Service:

- Research Data Management Programme
  - Delivered by Information Services
  - Supported by central funding
  - Phase 1: August 2012 to May 2015
  - Phase 2: June 2015 to July 2016
  - RDM Roadmap:
    http://www.ed.ac.uk/information-services/about/strategy-planning/rdm-roadmap

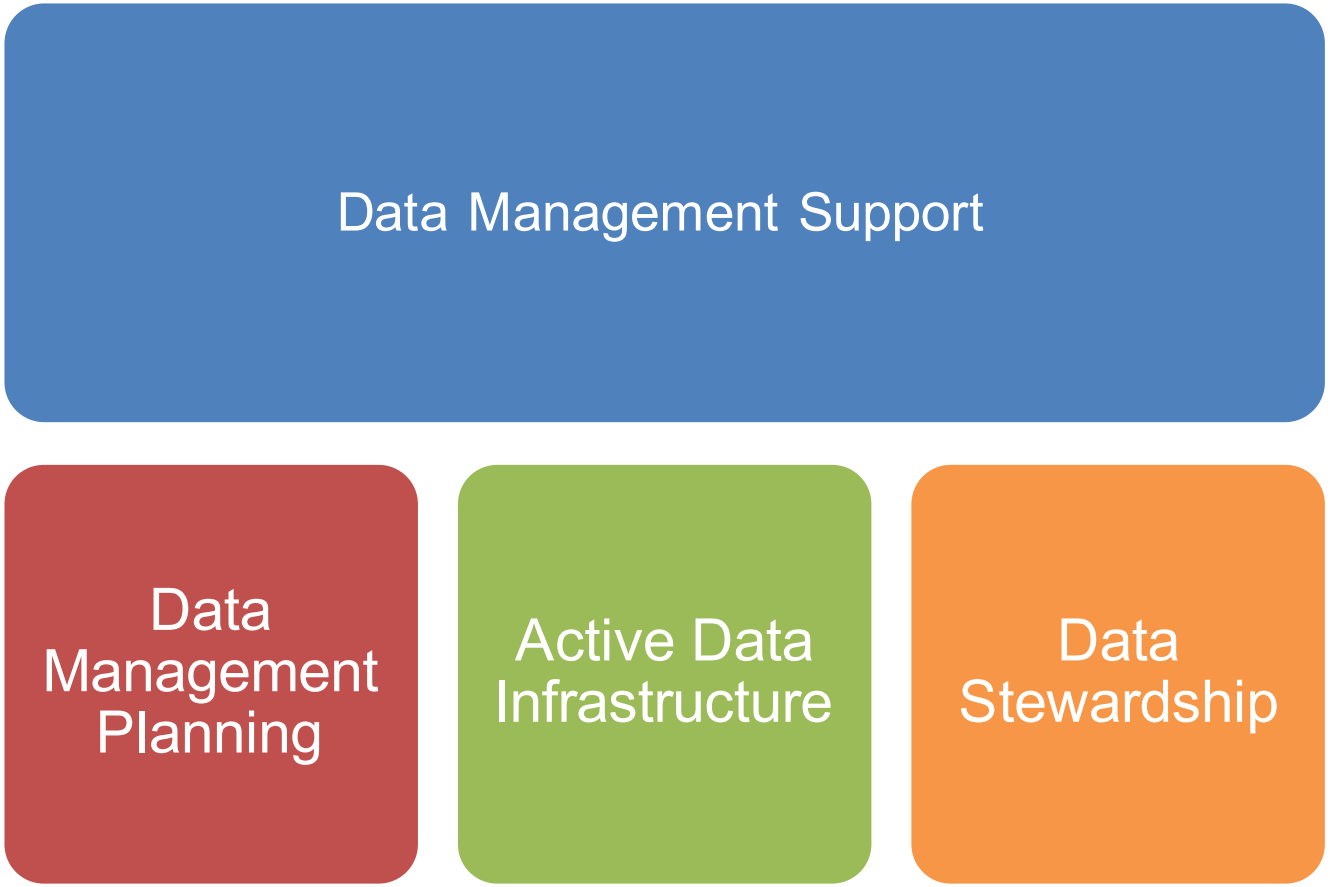
THE UNIVERSITY *of* EDINBURGH

Governance

- Academic-led steering group:
  - Chair: Professor Peter Clarke
  - Representatives from all Colleges
  - Professional services: Research Office, Library, IT
  - Meets quarterly

- 'Action group':
  - Chair: Stuart Lewis
  - Representatives from Library, IT, EDINA, Digital Curation Centre
  - Meets: Fortnightly

Research Data Management Services



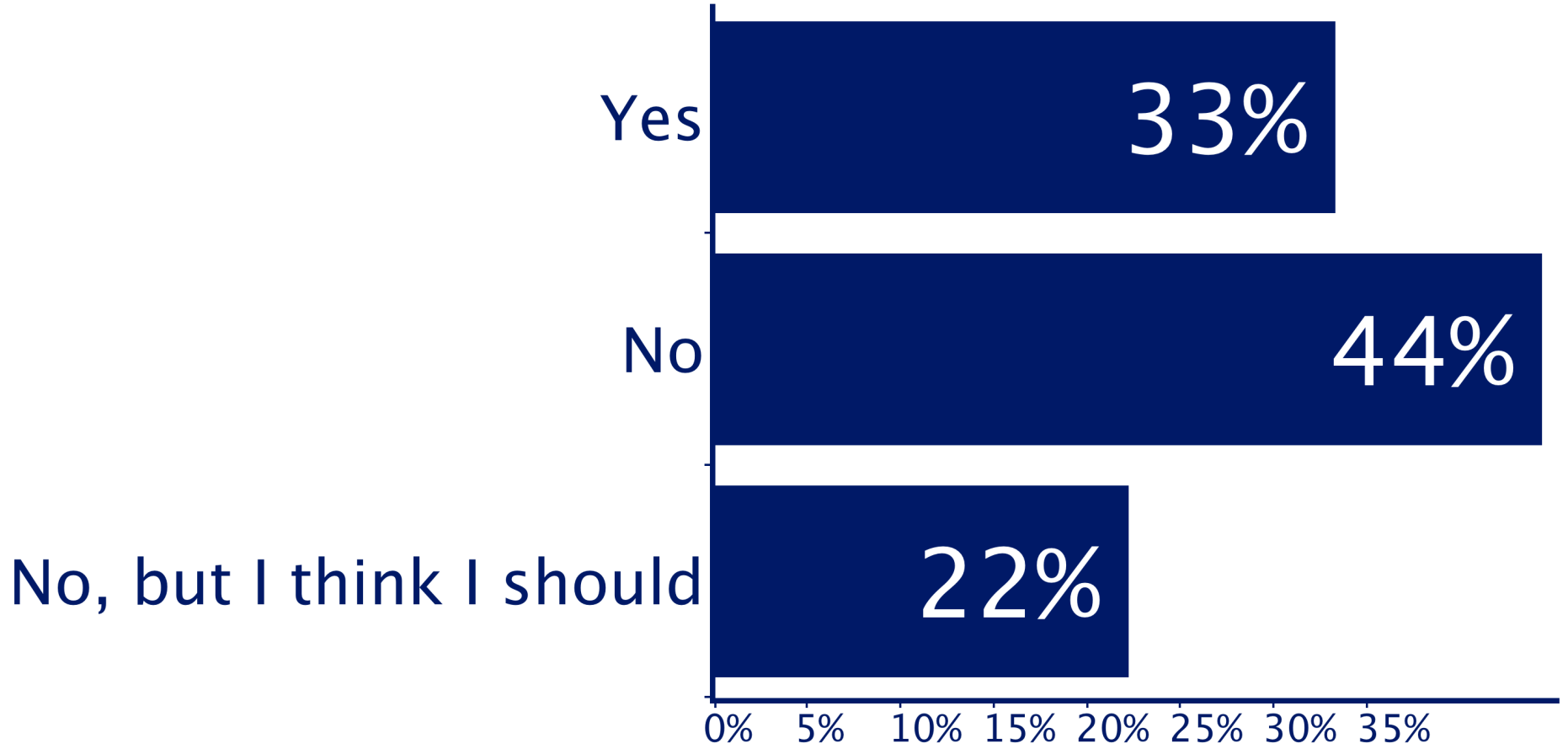
Data Management Planning

- DMPOnline National tool to create Data Management Plans
  - https://dmponline.dcc.ac.uk/

Sections	Questions
Data Capture	- Describe the types of data the research will generate - How will the data be documented? - How much data will be generated?
Data Management	- Where will the data be stored? - How will the data be backed-up? - Who will be responsible for this data management?
Integrity	- How will you quality assure your data?
Confidentiality and IPR	- How will you manage any ethical issues? - How will you manage copyright and Intellectual Property Rights (IPR) issues?
Retention & Preservation	- How long should the data be retained after the end of the research? - Which data should be retained after the research? - Identify the repository to which the data will be offered for deposit
Sharing & Publication	- Specify and justify which data will have value to others and should be shared - Indicate how data will be shared - Are any restrictions on data sharing required?

Have you ever created a Data Management Plan (DMP)?

Poll locked. Responses not accepted.



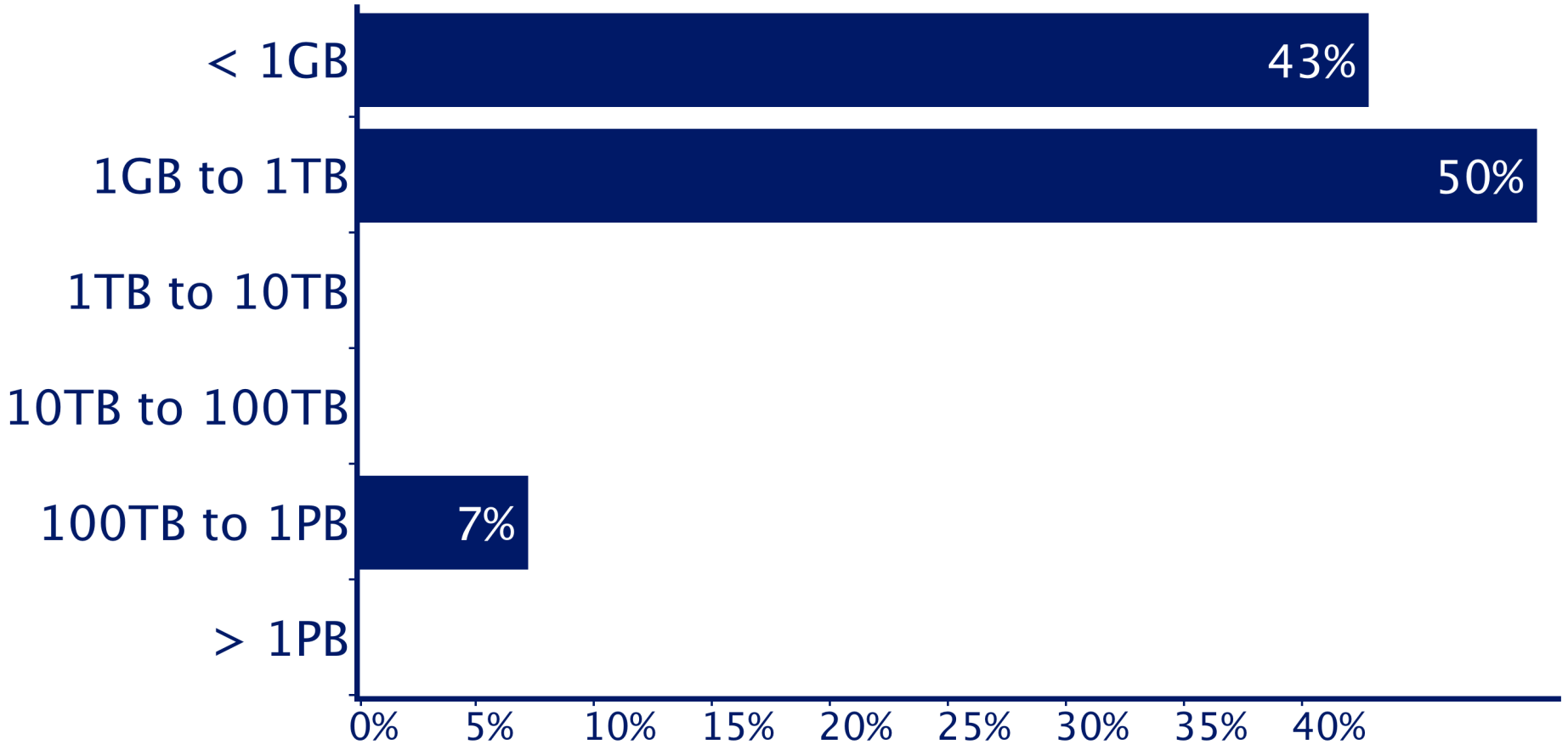
Active Data Infrastructure

- DataStore
  - 0.5 TB per person (PGRs upwards)
  - Personal allocation
  - 5TB per ‘project’
  - Extra can be purchased by grants @ £200 per TB per year
- DataSync
  - OwnCloud
    - Open Source DropBox-like web / sharing / sync system
  - Imminent launch
- Trusted Research Environment / Data Safe Haven

Edinburgh DataSync

How much data do you routinely manage?

Poll locked. Responses not accepted.



Active Data Infrastructure - collaboration

- Subversion
  - Source code control system
  - Allows software to be developed collaboratively
  - Possible move to GitLab (open source equivalent of GitHub)

- Wiki
  - Wiki for projects or teams
  - Atlassian Confluence

THE UNIVERSITY *of* EDINBURGH

Data Stewardship

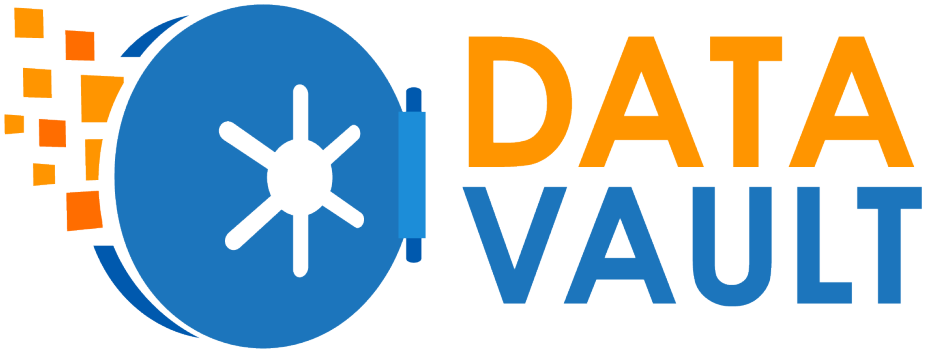
- PURE
  - Current Research Information System (CRIS)
  - Records research outputs (papers, grants, equipment, awards, datasets, etc)
  - Allows datasets to be described, and linked to if shared online

  - Person A, was awarded Grant B,
    which funded Equipment C,
    which created data D,
    which generated paper E

THE UNIVERSITY *of* EDINBURGH

Data Stewardship

- DataVault
  - Long term archival storage.
  - Move data from DataStore.
  - In development
    - With University of Manchester
    - Sponsored by Jisc
    - **http://datavaultplatform.org/**



THE UNIVERSITY *of* EDINBURGH

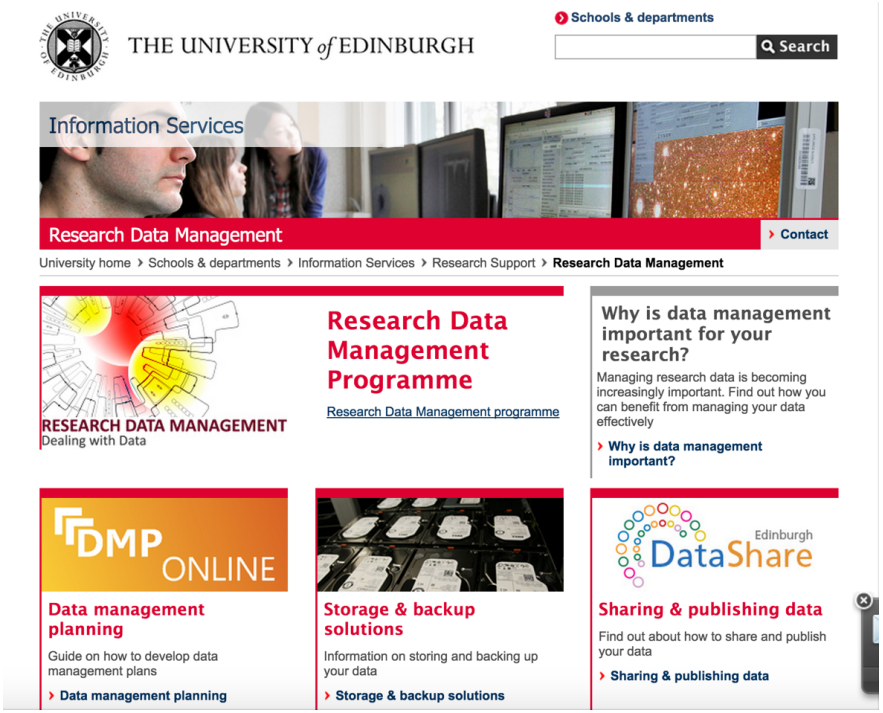
Data Stewardship

- DataShare
  - Online open data repository
  - Uses the DSpace open source repository platform
  - Creates DOIs for datasets
  - Data Seal of Approval (DSA)
  - http://datashare.is.ed.ac.uk/

THE UNIVERSITY *of* EDINBURGH

Data Management Support

- Awareness raising sessions
- Training courses
- On-demand support
- MANTRA online course
  - http://datalib.edina.ac.uk/mantra/
  - http://datalib.edina.ac.uk/mantra/libtraining.html



Moving towards Open Science

- Good Research Data Management is required for Open Science
  - Storing / Sharing
  - Reproducability
  - Transparancy

- Like Open Access, Open Science requires cultural change
  - Cultral change is slow
  - Compare to Open Access: 12 years, even through we're used to publishing
  - Try not to make 'openess' a barrier to RDM

THE UNIVERSITY *of* EDINBURGH

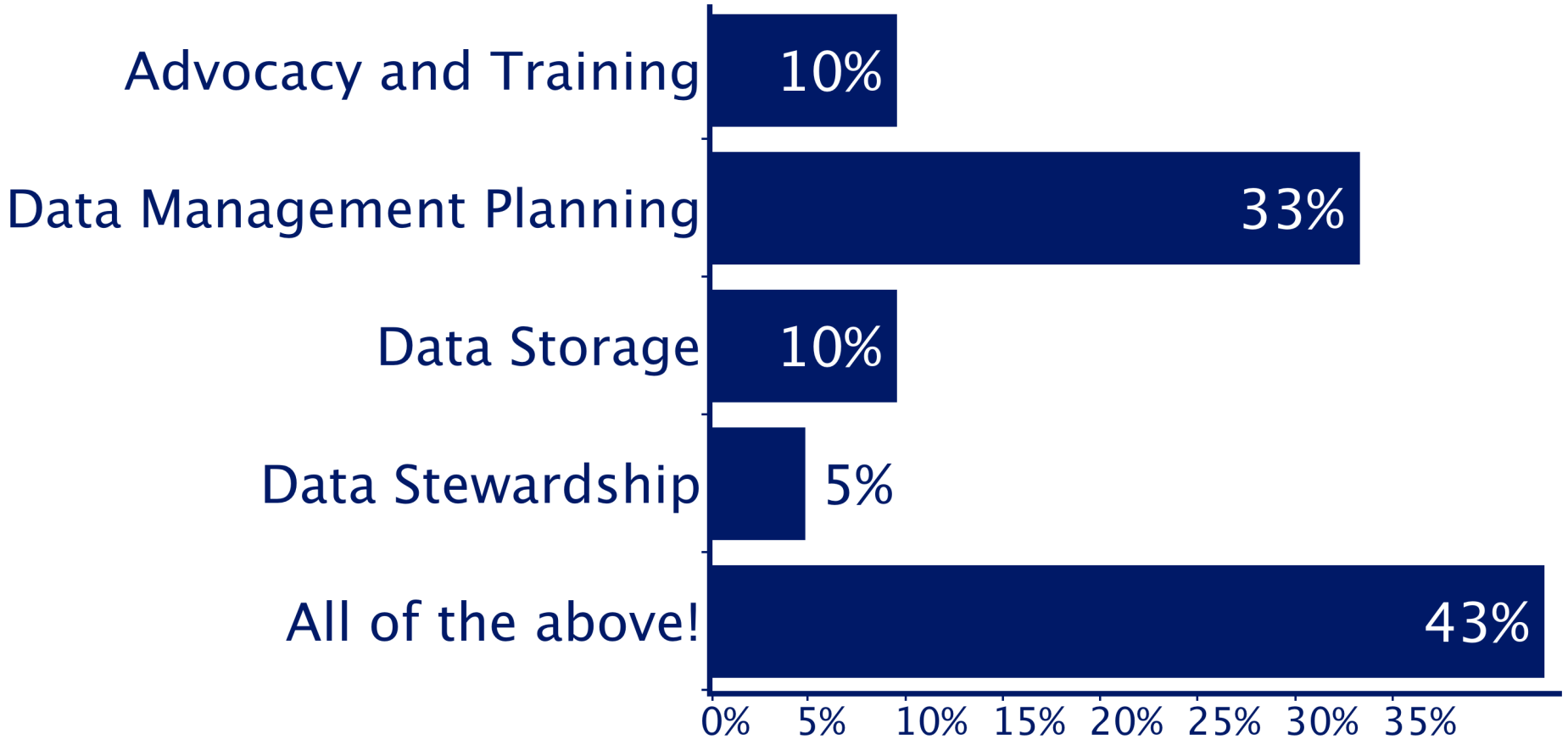
Challenges

- Funding – directly or indirectly from grants?
- Support staff skills – re-skilling and new sources
- Academic engagement – advocacy is slow, top-down and bottom-up
- Improving practice – new skills and approaches for research
- What / how much to keep? - starting to work with archivists
- Policies – standardise across funders
- Confusion – DataStore / DataSync / DataVault / DataShare
- Responsibilities – Who is responsible, for example security

THE UNIVERSITY *of* EDINBURGH

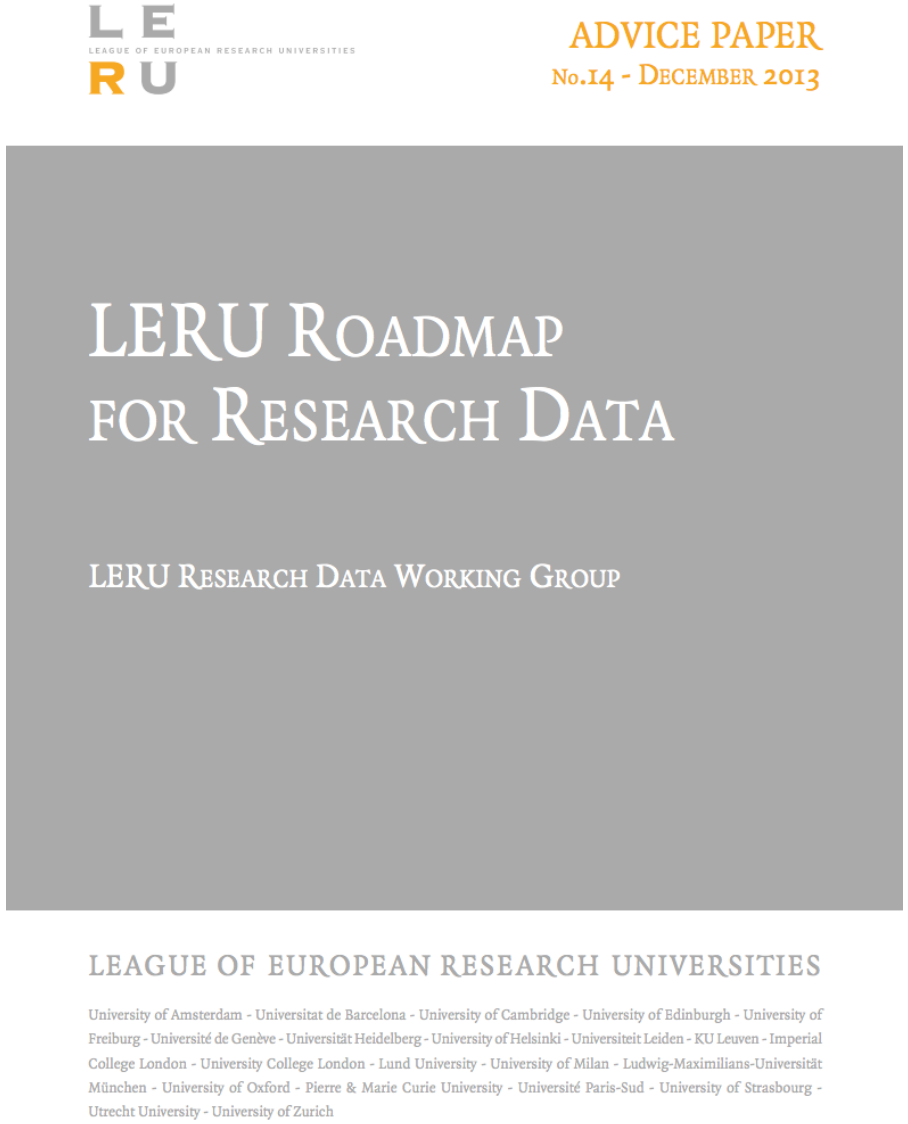
What RDM services would be most useful to start with?

Poll locked. Responses not accepted.



Useful reading

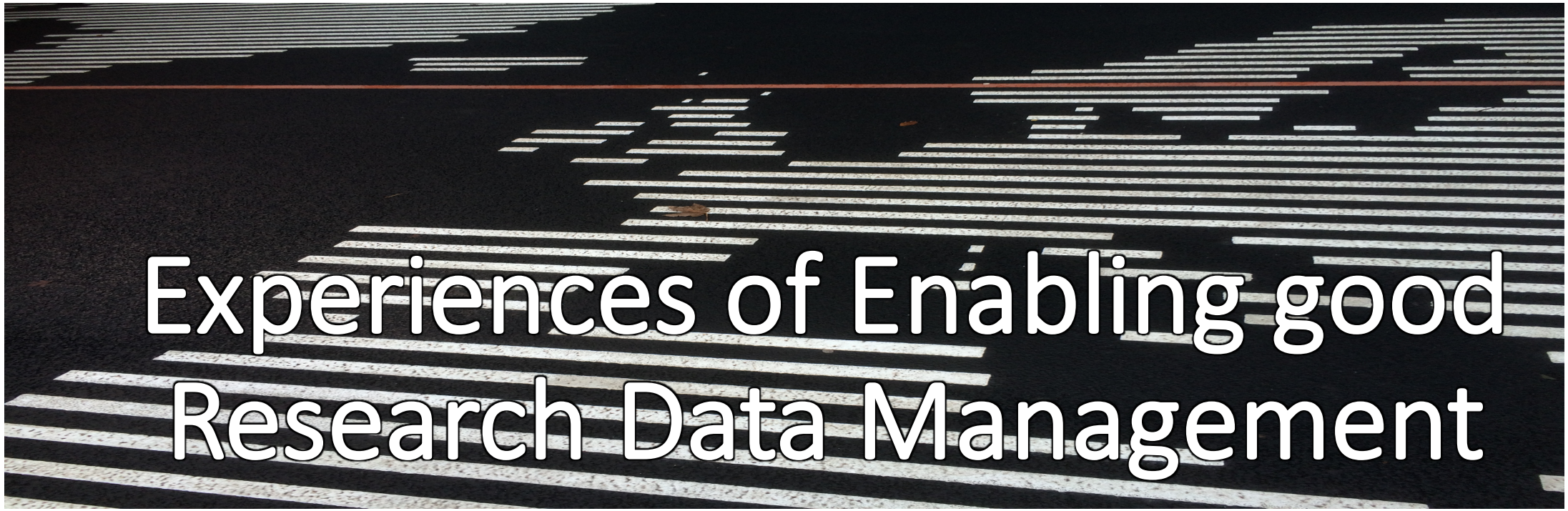
- LERU
  - LERU Roadmap for Research Data
  - http://www.leru.org/



Useful conference

- 11th International Digital Curation Conference (IDCC)
  - "Visible data, invisible infrastructure"
  - 22-25 February 2016, Amsterdam
  - "Research data management, and digital curation generally, is becoming a mainstream academic activity. Universities and research institutions are ramping up support and infrastructural provision for it. Funding bodies are strengthening their requirements for it. Visionary researchers and practitioners are tackling the remaining barriers."

http://www.dcc.ac.uk/events/idcc16



Experiences of Enabling good Research Data Management

stuart.lewis@ed.ac.uk

Þakka ykkur kærlega fyrir!

THE UNIVERSITY of EDINBURGH